Roberts, S. G., Sheard, C., Opie, C., Passmore, S., Jordan, F., & al., E. (2020). CHIELD: The causal hypotheses in Evolutionary Linguistics Database. *Journal of Language Evolution*, *5*(2), 101-120. https://doi.org/10.1093/jole/lzaa001

CHIELD:
The Causal Hypotheses in Evolutionary Linguistics Database

Seán G. Roberts
EXCD.LAB, Department of Anthropology and Archaeology, University of Bristol, 43 Woodland Rd, Bristol BS8 1TH, UK.

Anton Killin
School of Philosophy and Centre of Excellence for the Dynamics of Language, Australian National University, Canberra, ACT, Australia.
Department of Philosophy, Mount Allison University, Sackville, New Brunswick, E4L 1G9, Canada.

Angarika Deb
Department of Cognitive Science , Central European University, Oktober 6 street 7, 1st floor, Budapest, 1051, Hungary.

Catherine Sheard
School of Earth Sciences, University of Bristol, Life Sciences Building, 24 Tyndall Avenue, Bristol, UK.

Simon J. Greenhill
ARC Centre of Excellence for the Dynamics of Language, ANU College of Asia and the Pacific, Australian National University, Canberra 2600, Australia.
Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History, Jena 07743, Germany.

Kaius Sinnemäki
Department of Languages, University of Helsinki, 00014 University of Helsinki, Helsinki, Finland.

José Segovia-Martín
Cognitive Science and Language (CCiL), Universitat Autònoma de Barcelona, C/Montalegre, 6, 4ª planta, despatx 400908001 Barcelona.

Jonas Nölle
Centre for Language Evolution, The University of Edinburgh, Dugald Stewart Building, 3 Charles St, Edinburgh, EH8 9AD, UK.

Aleksandrs Berdicevskis
The Swedish language bank, University of Gothenburg, SE 405 30, Gothenburg, Sweden.

Archie Humphreys-Balkwill
EXCD.LAB, Department of Anthropology and Archaeology, University of Bristol, 43 Woodland Rd, Bristol BS8 1TH, UK.

Hannah Little

Science Communication Unit, Department of Applied Sciences, University of the West of England, Bristol, UK.

Christopher Opie
Department of Anthropology and Archaeology, University of Bristol, 43 Woodland Rd, Bristol BS8 1TH, UK.

Guillaume Jacques
CNRS-EHESS-INALCO, Centre de recherches linguistiques sur l'Asie orientale, 105 Boulevard Raspail 75006, Paris, France.

Lindell Bromham
Research School of Biology, Australian National University, 134 Linnaeus Way, Acton ACT 2601, Australia

Peeter Tinits
Department of Social Science, University of Tartu, Salme 1a–29, 50103, Tartu, Estonia.

Robert M. Ross
Department of Philosophy, Macquarie University, Level 2 North Australian Hearing Hub, NSW 2109, Australia.

Sean Lee
Graduate School of Asia-Pacific Studies, Waseda University, 1 Chome-21-1 Nishiwaseda, Shinjuku City, Tokyo 169-0051, Japan.

Emily Gasser
Linguistics Department, Swarthmore College, 500 College Avenue, Swarthmore, PA 19081, USA.

Jasmine Calladine
EXCD.LAB, Department of Anthropology and Archaeology, University of Bristol, 43 Woodland Rd, Bristol BS8 1TH, UK.

Matthew Spike
Centre for Language Evolution, The University of Edinburgh, Dugald Stewart Building, 3 Charles St, Edinburgh, EH8 9AD, UK.

Stephen Francis Mann
School of Philosophy and ARC Centre of Excellence for the Dynamics of Language, Australian National University, H.C. Coombs Building, ACT 2601, Australia.

Olena Shcherbakova
Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History, Jena 07743, Germany.

Ruth Singer

School of Languages and Linguistics, University of Melbourne, Babel (Building 139), Parkville 3010, VIC, Australia.

Shuya Zhang
CNRS-EHESS-INALCO, Centre de recherches linguistiques sur l'Asie orientale, 105 Boulevard Raspail 75006, Paris, France.

Antonio Benítez-Burraco
Department of Spanish, Linguistics, and Theory of Literature (Linguistics). University of Seville. Palos de la Frontera, 41004, Seville, Spain.

Christian Kliesch
Department of Psychology, Lancaster University, Lancaster, LA1 4YF, UK.
Max Planck Institute for Human Cognitive and Brain Sciences. Stephanstraße 1a, 04103 Leipzig, Germany.

Ewan Thomas-Colquhoun
EXCD.LAB, Department of Anthropology and Archaeology, University of Bristol, 43 Woodland Rd, Bristol BS8 1TH, UK.

Hedvig Skirgård
School of Culture, History and Language, Australian National University, Canberra, ACT, Australia.

Monica Tamariz
Psychology, Heriot-Watt University, Edinburgh, EH14 4AS, UK.

Sam Passmore
EXCD.LAB, Department of Anthropology and Archaeology, University of Bristol, 43 Woodland Rd, Bristol BS8 1TH, UK.

Thomas Pellard
CNRS-EHESS-INALCO, Centre de recherches linguistiques sur l'Asie orientale, 105 Boulevard Raspail 75006, Paris, France.

Fiona Jordan
EXCD.LAB, Department of Anthropology and Archaeology, University of Bristol, 43 Woodland Rd, Bristol BS8 1TH, UK.

Abstract

Language is one of the most complex of human traits. There are many hypotheses about how it originated, what factors shaped its diversity, and what ongoing processes drive how it changes. We present the Causal Hypotheses in Evolutionary Linguistics Database (CHIELD, https://chield.excd.org/), a tool for expressing, exploring, and evaluating hypotheses. It allows researchers to integrate multiple theories into a coherent narrative, helping to design future

research. We present design goals, a formal specification, and an implementation for this database. Source code is freely available for other fields to take advantage of this tool. Some initial results are presented, including identifying conflicts in theories about gossip and ritual, comparing hypotheses relating population size and morphological complexity, and an author relation network.

1 Introduction

Evolutionary linguistics is a field that uses evolutionary principles to explain the origins of complex communication systems, as well as the similarities and differences between them (see e.g., Knight, Studdert-Kennedy & Hurford, 2000; Wray, 2002; Botha, 2003; Christiansen & Kirby, 2003; Hurford, 2007; Kinsella, 2009; Fitch, 2010; Berwick & Chomsky, 2016; Progovac, 2019). Scott-Phillips & Kirby (2010) identified four phases in human language evolution that are studied within this field:

- *preadaptation* - the preconditions for a language ability, often related to genetic evolution (e.g., Lieberman, 1984; Corballis, 1999; Hurford, 2003; Slocombe & Zuberhühler, 2005; Cheney & Seyfarth, 2005; Fehér, 2017; Vernes, 2017).
- *co-evolution* - how the first human communication systems and these preadaptations evolved together (e.g., Deacon, 1997; Dor, Knight & Lewis, 2014; Berwick & Chomsky, 2013; Pakendorf, 2014; Vigliocco, Perniss & Vinson, 2014; Power, Finnegan & Callan, 2016; Falk, 2016).
- *cultural evolution* - the initial emergence of new linguistic structures (e.g., Nowak & Krakauer, 1999; Tallerman, 2007; Smith & Kirby, 2008; Culbertson & Newport, 2014; Progovac, 2015; Kempe, Gauvrit & Forsyth, 2015; Tamariz & Kirby, 2016; Goldin-Meadow & Yang, 2016; Piantadosi & Fedorenko, 2017).
- *language change* - the ongoing change in languages (e.g., Mufwene, 2001; Ritt, 2004; Croft, 2008; Sampson & Trudgill, 2009; Dunn et al., 2011; Gavin et al. 2013; Majid, Jordan & Dunn, 2015; Bowern, 2015; Bybee, 2015; Coelho et al., 2019).

There have been many exciting developments in recent decades, making it perhaps possible to join them into larger theories. However, synthesis has become difficult as there now exists a mountain of theories and evidence, in increasingly specialised sub-fields. Jim Hurford once moderated a discussion between four plenary speakers who had presented four different theories of language evolution. His first question was “What do you disagree about?”. Nobody had a reply. This showed that, although the theories were internally consistent, they weren’t connected to each other. This characterises a problem in many fields - it is possible to have nearly as many theories as there are researchers, and debate is often limited to dogmatic acceptance or complete rejection of these theories, rather than trying to systematically compare, evaluate, and synthesise.

Progress in evolutionary linguistics will be made by working towards building a chain of causal links that join theories together. Since there are many aspects of language evolution that cannot be tested directly, each link should be tested with multiple methods and sources of data - a ‘robust’ approach (Irvine, Roberts & Kirby, 2015; Roberts, 2018). In order to combine these different strands of evidence (experiments, models, simulations, comparative work), researchers must coherently express how these results relate to each other and to the real

world (Vogt & De Boer, 2010). There have been previous calls for this kind of approach; for example, Zuidema & de Boer (2010, 2013) suggest that computational modelling should aim for greater 'model parallelisation' (qualitatively comparing models against each other) and greater 'model sequencing' (building chains of models that feed into each other). However, practical solutions are difficult to produce, due to challenges that relate to expression, exploration, evaluation and extension.

Expression: Expressing complex causal hypotheses in prose is difficult, and many hypotheses are underspecified. Without deep knowledge of a theory's particular sub-field, it is often difficult to identify the assumptions and claims of a theory. In addition, different sub-fields may use different terms to refer to very similar concepts, or use the same term to refer to different concepts. A classic example of this is the word "language" itself, which can be interpreted as anything from relating to human communication to only a specific syntactic ability. All this can lead to researchers talking past each other, along with a general lack of connection between sub-fields. How can we better express hypotheses to avoid these problems?

Exploration: Evolutionary linguistics now includes many subfields within linguistics (Bergmann & Dale, 2016; Berwick & Chomsky, 2016), and also relates to other larger interdisciplinary fields of research, such as learning or cooperation (Kirby & Christiansen, 2003; Progovac, 2019). The methods are diverse, ranging from molecular genetics (e.g., Enard et al., 2002; Hitchcock, Paracchini & Gardner, 2019), to archaeology (e.g., Noble & Davidson, 1996; Currie & Killin, 2019), to computational simulation (Steels, 1997; Cangelosi & Parisi, 2012; Jon-And & Aguilar, 2019). It is therefore increasingly hard to keep up to date with all the developments in the field. How can we make theories searchable so that researchers can access work from other fields that relates to their own? Furthermore, how can we formally relate these hypotheses to each other, in order to find similarities and differences?

Evaluation: After relating theories to each other, how do we then evaluate them? How do we identify the claims that are supported by evidence, and those that require further investigation? How can we get an overview of research conducted on a topic? Studies are getting bigger, and there is more emphasis on large-scale tests that can evaluate multiple competing models. Systematically collecting these hypotheses, storing them and converting them into statistical models is hard (Bareinboim & Pearl, 2016).

Extension: Language evolution is a field that has inspired much debate, and even reaching a consensus on interpretations of hypotheses is difficult. How can we support researchers in the continuing process of refining and extending them? How can we ensure that the tools to do this will be useful into the future?

One possible solution is to harness the power of causal graphs. A causal graph is a graphical tool which breaks a complex hypothesis into individual causal links. We present the Causal Hypotheses in Evolutionary Linguistics Database (CHIELD, pronounced like "shield", https://chield.excd.org/), a database of hypotheses expressed as causal graphs. It allows users to apply computational search and visualisation methods, in order to express, explore and evaluate hypotheses. This paper describes the design and presents three case studies to demonstrate its functionality. Case Study 1 demonstrates how CHIELD can be used to explore connections between theories. Case Study 2 attempts to evaluate competing explanations of

the connection between population size and morphological complexity, and demonstrates some issues with vocabulary. Of course, problems such as converging on the same vocabulary requires more than a database to solve, but CHIELD may at least provide a space for spotting potential areas of disagreement. Case Study 3 demonstrates extended uses such as constructing networks of authors working on the same topics.

To be clear, the aim is not to build a list of theories that have been accepted by the scientific community as "correct", that are somehow more "prestigious" or supposedly have no counterevidence. It is, of course, very difficult to prove a causal effect as being distinct from a correlation. However, at the heart of any research trying to explain a phenomenon is an idea about some kind of causal relationship. The aim therefore is to faithfully represent these ideas such that researchers can plan future work. Furthermore, the aim of this database is not to highlight one view over another, but to simply present them on an accessible platform. Its aim is description, not prescription. Finally, the database aims to be editable and maintainable into the future. We hope that future studies in language evolution will be enhanced by the insights provided by causal graphs.

2 Causal inference and causal graphs

Causal inference is an approach to thinking about causality (Pearl, 2000; Pearl & Mackenzie, 2018; Rohrer, 2018) that uses graphical tools - causal graphs - to depict models of causal processes. Causal graphs consist of nodes that represent measurable quantities or concepts, and arrows which represent the causal influences between these nodes. For example, consider why we might see more shirt stains on hot days. Prof. Whippy suggests a causal explanation: “High temperatures cause more ice-cream consumption, and more ice-cream consumption leads to more shirt stains". So we can draw the following causal graph:

Temperature → Ice cream consumption → Shirt stains

This causal graph has three measurable quantities (nodes e.g., “Temperature”) and two causal links. The links are interpreted according to an interventionist interpretation of causality (see e.g., Woodward 2003, 2016). For example, the first causal link states that if one were to “intervene” by changing the temperature sufficiently, then the amount of ice cream consumption would also change (and the second link states that if one were to increase the amount of ice cream consumption sufficiently, then the number of shirt stains would increase). That is, it makes a counterfactual claim about a possible world where the measurable quantities were different. Interventionist causality is often interpreted in experimental terms: if one were to experimentally manipulate the temperature, then there would be a change in ice cream consumption. Relationships encoded by the causal arrow could be fully deterministic or just probabilistic, and could be between continuous, discrete or categorical variables. This is a widely used approach for which many helpful tools are available.

Note that the graph above also makes some more claims. Firstly, there is no link from ice cream consumption to temperature, which is interpreted as there being no causal effect of ice cream consumption on temperature. That is, if we were to intervene by forcing people to consume ice cream, then the temperature would not increase. Secondly, since there is no

direct causal link between the first and the third node, the number of shirt stains is causally independent of the temperature such that temperature only influences shirt stains via ice cream consumption. This hypothesis could be experimentally tested, for example by raising the temperature (maybe within a shopping centre) and simultaneously banning the sale of ice cream, to see if the number of shirt stains decreases.

A key thing to understand about causal graphs is that they do not necessarily reflect what has been proven to be true, but instead reflect a particular researcher's hypothesis. In other words, causal graphs represent *ideas* about how the world works, they are not proof that the world really is like that. There are many other possible hypotheses (and causal chains) to explain the same phenomenon. For example, Prof. Whippy suggested that higher temperature causes more ice cream consumption, which in turn causes more shirt stains. Let's imagine that, in another paper, Prof. Gelato attacks Prof. Whippy's hypothesis and suggests a different explanation, namely that ice cream consumption has no effect on shirt stains, and instead seeing a shirt stain reminds people of ice cream and so they seek it out. In yet another paper, Prof. Sorbet studies climate change and suggests that refrigerated ice cream vans are contributing to greenhouse gases, therefore affecting the temperature. We can draw the three hypotheses in a single graph (Figure 1).

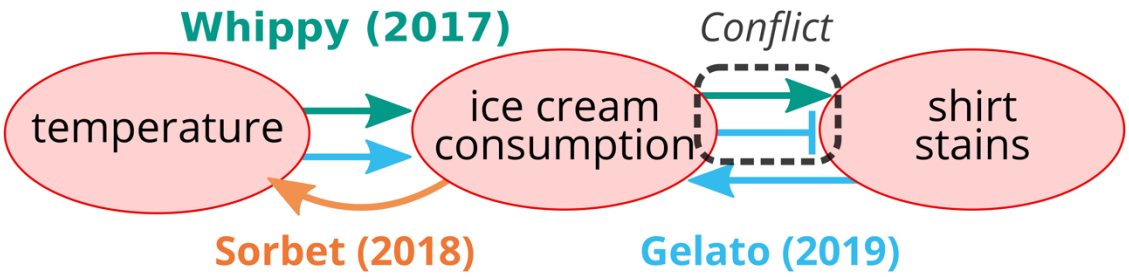

Figure 1: Causal graph showing three hypotheses of the relationship between temperature, ice cream consumption and shirt stains. Nodes represent measurable quantities, and lines between them represent hypothesised causal links. Lines with arrows reflect causal effects and the line with a bar head is interpreted as 'no causal effect'. Lines are coloured according to their source publication. The gray dotted rectangle highlights a conflict between two theories.

By representing multiple hypotheses on a single causal graph, we can identify relationships between them. For example, Profs Whippy and Gelato agree on the effect of temperature on ice cream consumption. Furthermore, Prof. Sorbet's climate change mechanism is not necessarily in conflict with the effect of temperature on ice cream consumption. Both could be operating to create a feedback loop. However, Prof. Whippy and Prof. Gelato disagree on the relationship between ice cream consumption and shirt stains. Intuitively, Prof. Gelato's theory seems much less likely to be true, but the causal graph is still a valid representation of Prof. Gelato's theory. Drawing out the two theories has therefore shown where the conflict lies, and further suggests a future empirical test: Prof. Gelato would predict that staining people's shirts would cause ice cream consumption to increase, while Prof. Whippy would predict that it would make no difference.

As this example shows, expressing different hypotheses as causal graphs allows researchers to express, explore, evaluate, and extend the relationships between them (see Höfler et al., 2018). Constructing a causal graph is also an excellent way to identify underspecified mechanisms. Causal graphs can help readers to understand papers (Cao, Sun & Zhuge, 2018; Easterday, Aleven & Scheines, 2007; Höfler et al., 2018; although Tubau, 2008, suggests that they may not aid reasoning in certain domains) and facilitate debate (Easterday, et al., 2009). Causal graphs also help communicate theories indirectly by providing a roadmap for writing (e.g., authors can check if they have provided justification or evidence for each causal link, or identify arguments that are tangential to the central claim, see section 5.2 below).

Expression of theories is an important part of teaching, and a unified approach to expressing causal hypotheses in language evolution could improve student understanding. For example, students could point to a particular causal link that they do not understand, and the database could give them a quote from the paper. Causal graphs also aid understanding in high-school students (Hsu, Van Dyke, Chen & Smith, 2015).

Of course, carrying out the experimental manipulations mentioned above might be impractical. In this case, researchers might depend on converging evidence from multiple sources, analogies with simulations or 'natural experiments' (see e.g., Steels, 1997; Pyers et al., 2010; Morgan, 2013; De Boer & Verhoef, 2012; Irvine, Roberts & Kirby, 2013). For example, in the case above, researchers could seek locations where the temperature does not change, or where the temperature varies but there is no ice-cream. However, many recent developments in causal inference have allowed researchers to estimate causal effects with purely observational data, even in cases where researchers cannot directly measure a variable of interest (Pearl, 2000; Bareinboim & Pearl, 2012).

Another advantage is in identifying conditioning sets: which variables to control for in an analysis. Standard methods about how to choose control variables are often vague (Pearl & Mackenzie, 2018), and many assume that controlling for more variables makes the central test more robust. However, controlling for some variables can create spurious correlations due to *colliders*. A collider is a node on a causal path with two causal links (arrows) pointing into it (see supporting materials S5). In the example above, Prof. Gelato predicts that 'ice cream consumption' is a collider along the path involving temperature and shirt stains. Gelato would predict that temperature and shirt stains both contribute to ice cream consumption. If this were true, then temperature and shirt stains should be uncorrelated, except when controlling for ice cream consumption, at which point they would become correlated (see e.g., Elwert 2013). This property of colliders means that spurious correlations can sometimes be introduced by statistical controls (see Elwert & Winship, 2014; Ding & Miratrix, 2015; Westfall & Yarkoni, 2016; Rohrer, 2018; Middleton et al., 2016; York, 2018). Drawing a causal graph helps to identify these cases, and therefore helps make effective decisions in study design and analysis. These benefits make causal graphs a very powerful tool for research in the social sciences. However, causal graphs are not substitutes for careful thought and design. They are tools for helping researchers do their job. CHIELD aims to make these tools more accessible.

Some recent developments in causal inference theory go further by attempting to estimate the most likely causal graph directly from observational data on large number of variables (Kalisch

et al., 2012; Hauser & Bühlmann, 2012; see Heinze-Deml, Maathuis & Meinshausen, 2018, for applications in linguistics see Roberts & Winters, 2013; Baayen, Milin & Ramscar, 2016; Blasi & Roberts, 2017). Analysing the entire causal network, rather than just explaining the variation in a single target variable, makes it possible to study much more complex relationships, and to form a more comprehensive picture of complex systems such as those found in social sciences. However, there are many possible ways of applying this method (parameters of the algorithm, treatment of the data, assumptions of the statistical tests), and these can lead to big differences in the results and interpretation. What is needed is a way of compiling and organising existing knowledge about causal effects within one domain, in order to evaluate the automatically derived causal graph. CHIELD aims to provide this prior knowledge in order to protect researchers from making hasty post-hoc hypotheses from the output of automatic methods.

3 Design of the database

We reviewed various existing applications (see supporting materials S1) to check whether they address the problems discussed above. There appears to be no single tool that allows researchers to express hypotheses (visually), explore the way they interact, and evaluate them. CHIELD aims to fill this gap. In this section, we cover the design of CHIELD. The principle is to store data in a simple spreadsheet format, with scripts that convert them to a database that can be accessed and edited through a customised web interface.

3.1 Expression

The aim of CHIELD is to allow researchers to express causal graphs for a particular theory, in a simple and intuitive format. CHIELD is based around ‘documents’ (usually a peer-reviewed, published paper), each of which have three associated files: bibliographic information stored as a bibtex file, a simple text file listing who contributed the data, and a spreadsheet file which stores information about the causal links in the document. These files are kept together in a folder, which is named after the unique bibtex key for the document. In order to make finding particular files easier, we borrow the design of Glottolog: document folders are sorted into parent folders for each year of publication, and then each year is sorted into parent folders for each decade of publication (see https://github.com/CHIELDOnline/CHIELD/tree/master/data/tree/documents). For example, the files for Dunbar’s (2004) paper on gossip is stored in ‘documents/2000s/2004/dunbar2004gossip/’. This organisation is largely for the convenience of the database developer. For most users, the online interface for coding papers will automatically create files in the correct location, and documents can be searched using a user-friendly interface.

Within the causal links spreadsheet, a single observation (a single row) represents a single causal link between two variables. The causal link has a number of associated properties, shown in table 1. The types of causal relation are shown in table 2. While a standard causal graph does not encode the direction of the relationship, this information can be encoded in

CHIELD via the “Cor” (correlation) field, including positive, negative and non-monotonic relationships.

The format is easiest to explain with an example (see table 3). Lupyan & Dale (2010) suggested a link between the number of speakers a language has, and the language’s morphological complexity. In their paper, they hypothesise that: “Speakers of languages [with large numbers of speakers] are more likely to use the language to speak to outsiders—individuals from different ethnic and/or linguistic backgrounds.” This could be coded as the first row of table 3. They hypothesise a causal effect (“>”) and they support this with references to the literature (type = “review”). The main quantitative results from the paper demonstrate a negative correlation between population size and morphological complexity. This can be coded separately as a “statistical” type of support. See section 5.3 or the entry in CHIELD (https://chield.excd.org/document.html?key=lupyan2010language) for more information. More detailed specifications for each field are provided in the supporting materials.

Defining hierarchies of variables is possible by using a colon character in the variable name. For example, in a paper on morphological complexity, Nettle (2012) distinguishes “paradigmatic complexity” from “syntagmatic complexity”. Ideally, these concepts should link to the “morphological complexity” node of other hypotheses while maintaining their specificity. Therefore, a coder might use “morphological complexity: paradigmatic” and “morphological complexity: syntagmatic”. Various settings in the visualisation allow users to switch from connecting nodes only if their full labels match, and connecting them if their higher-order category matches.

It would also be possible to allow empirical measures of causal strength to be stored for each causal link, which are important for causal studies in fields like medicine. However, they are not part of the core design goals here, since many studies of evolutionary linguistics do not estimate these kinds of measures directly, but use experiments, models or case studies to make analogies. The most important design feature here is a low barrier to adding data, and causal strength estimates may not be very helpful for many researchers in evolutionary linguistics. Having said this, it would be easy to add this kind of information to a future version of the database.

3.1 Exploration

The web interface for CHIELD includes several ways of searching the data. PHP scripts fetch data from the compiled SQLite version of the database and serve them up into a dynamically searchable tabulated format (using the *Datatables* library, https://datatables.net). For example, users can view a list of documents, and search by title, author or year. Each document links to a document page, which displays the bibliographic information, the list of causal links (with the specification information above) and a list of other documents that have variables in common (found using live queries of the SQLite database). CHIELD also includes an interactive visualisation of the causal graph using the *vis.js* javascript library (http://visjs.org). Users can view tables which list all the causal links or all variables. These link to similar pages which give overviews of the variable (e.g., all causal links involving

‘linguistic diversity’). These search features make it easy to find documents or causal links, and to explore links between documents.

Collaboration is an important part of language evolution research (see Bergmann & Dale, 2016; Youngblood & Lahti, 2018). There are now various approaches for automatically suggesting collaborations between authors based on network models (e.g., Lopes et al., 2010; Xu et al., 2010; Yan & Guns, 2014; Guns & Rousseau, 2014; Kong et al., 2017). By combining the bibliographic information with the causal links data, CHIELD may be able to find connections between authors beyond co-authorship. Users can search a list of authors, with links to author pages showing all documents and causal links by an author, a list of their co-authors and a list of potential collaborators. The potential collaborators are defined following a method similar to the AXON database (see supporting materials S1): find authors whose causal graphs overlap (have nodes in common), but who have not published any papers together.

CHIELD includes an “Explore” mode, where users can load multiple causal graphs into a single visualisation. This includes various tools:

- An interface for combining multiple causal graphs into a single visualisation.
- Interactive manipulation, allowing adding a node from the database, adding all nodes from a particular document, or removing nodes or edges from the current visualisation.
- Expand links from the currently selected node (display all causal links that connect to a particular node).
- Find evidence from other documents for the currently displayed links.
- Display sub-variables under a single higher-order variable
- Find causal pathways between two given variables, in order to discover alternative hypotheses.
- Exporting the causal graphs as ‘dot’ format files. These can be used in visualisation tools, like GraphViz (Ellson et al., 2004) or Gephi (Bastian, Heymann & Jacomy, 2009) to produce images for use in publications (e.g. pdf, svg).

These tools can help a researcher find “upstream” explanations that feed into their own hypothesis, and also see how their hypothesis can generate predictions for other “downstream” work.

CHIELD finds causal pathways using a variant of Dijkstra's algorithm (Dijkstra, 1959), which finds all possible paths between two nodes. This can become complicated if there are loops or multiple connections between variables. However, the algorithm does not need to find all possible paths, only the set of nodes along all possible paths (the SQL query will find all paths connecting these nodes later); the algorithm only follows each edge once. Another issue is which types of causal connection should be considered in the search. Considering only standard causal links (“>”) can lead to missing some connections, but including all links can lead to very large causal networks which are not useful. The current implementation, therefore, only considers the following connections: ">", ">>", "<=>" and “=~". In the future, this search process could be customisable.

3.2 Evaluation

The explore mode allows control over the visualisation in order to support evaluation and effective design decisions for future research. Parameters of the causal graph can be manipulated (e.g., to display hierarchical layout or a dynamic 'spring' layout). For example, the colours of edges and node positions can be manipulated to reflect the source publication, type of evidence, type of causal effect or direction of correlation (communicated through a hideable legend). The colour schemes are designed to be distinguishable by colourblind users.

The explore mode also includes a tool for highlighting differences between hypotheses. If causal links from more than one document are loaded, an algorithm finds edges where the two documents disagree. At the moment, this only involves cases where one document claims that there is a causal connection (">", ">>", "<=>") and the other claims that there is no causal connection ("/>"), but this could be expanded in the future. If conflicts are detected, the relevant edges are highlighted and the visualisation zooms in to display them.

The final causal graph can be exported to the DAGitty web interface (Textor, Hardt & Knüppel, 2011) or as model definition code for the R package *phylopath* (von Hardenberg & Gonzalez-Voyer, 2013; van der Bijl, 2018), which performs phylogenetic path analysis on multiple competing models (each document is listed as a different model). The full database is openly available in a range of formats, for further manipulation by statistical software.

3.3 Extension

Constructing a comprehensive database of theories of language evolution is not a feasible task for a single person. Consequently, CHIELD has been designed to be extendable by large numbers of contributors. This makes it important to be able to curate contributions and keep track of changes. It is not expected that everyone will agree on interpretations, but these issues are worth debating, and the database should include space for discussion and revisions. If the database is to be useful in the long-term, there are also the practical questions about how best to store the data. Moreover, we want the basic application to be extendable to other fields and by other developers. Therefore, the design goals here are:

- Integration with version control software for keeping track of changes.
- Tools for discussion and curation.
- Simple file formats that can be easily edited.
- Non-proprietary data formats for longevity.
- Open source code for extension to other fields.

CHIELD is integrated with Git and GitHub for keeping track of changes, and for managing public contributions, issues and discussion (https://github.com/CHIELDOnline/CHIELD). This also provides open access to the source code if other fields of research want to develop their own version. The file formats (bib, txt and csv) are open-source, non-proprietary and simple to edit. All processing scripts and external libraries are free and open source.

Data can be added to CHIELD by coders through the GitHub repository, but the website also includes a simpler customised interface for adding data (using the javascript library *js-grid*, http://js-grid.com/). This guides coders through the process of contributing a document to

CHIELD, including how to draw the causal graph visually or to upload a causal links template spreadsheet from their computer (more information at https://chield.excd.org/Help_AddingData.html). The interface suggests existing variable names for the coder to re-use, in order to maximise convergence on variable names.

Once the coder has entered their data, a script creates the standard file formats for CHIELD and sends them to the GitHub repository (it creates a new branch, commits the new files to it, and then makes a pull request to merge these into the main branch). This sends an alert to the owner of the repository who can review the contribution and accept it into the database. The web manager can now pull these changes down to the server, and run the compiling script that creates an updated SQL database and deploys it to the website. This system is perhaps one of the more successful parts of the database design that allows anyone to contribute without needing to know how to use version control software. This system also takes advantage of the existing tools of GitHub, including user logins, bug reports, discussion threads and tracking changes which reduces the development load. Finally, this system also allows rapid turnover: after the coder has submitted their data, the process of adding it to the database and updating the website takes just a few minutes.

3.3.1 Scope of data

The condition for entry into the database is that the document proposes a hypothesis with causal claims that relates to some part of the evolution of communication and that it is published in a peer-reviewed publication (e.g., journal paper, peer-reviewed conference proceedings, book chapter). This could be either biological or cultural evolution (or both) at any evolutionary stage. As a rule of thumb, the document should relate directly to language, or be “one link away” from a relevant language document. Entry into the database does not mean that the hypothesis is correct or widely accepted, or even empirically supported. The aim is not that the database be a single theory of the evolution of communication, but a reflection of the whole field. Potential sources for documents include the Language Evolution and Computation Bibliography (https://langev.com, including over 2,500 papers), the *Journal of Language Evolution*, the *Journal of Interaction Studies*, *EvoLang* conferences, and so on.

3.3.2 Moderating contributions

CHIELD aims to be open and editable. Anyone can contribute documents to CHIELD or edit any existing documents (as long as they have a free GitHub account). This necessitates moderation for quality control. This is done through GitHub, with a central administrator being able to review each contribution or edit (as a pull request) before it is added to the database. An advantage of this system is that it creates a record of every change to the database. The document pages of the website include buttons for raising issues with the coding (through a pre-filled GitHub issues page), which may be taken up by other users or an administrator. There are various administrator tools for aggregate tasks, like replacing all occurrences of a particular variable label with another.

4 Results

The CHIELD website is live (http://chield.excd.org/) and is updated continuously (https://github.com/CHIELDOnline/CHIELD). The results in this paper are based on version 1.1. In general, we found coding of causal graphs challenging but productive. We found that a paper might take between 10 minutes and an hour to code, depending on the complexity of the theory, the methods used and the coder's familiarity with the subject. Version 1.1 of CHIELD includes 400 documents and 3,406 causal links between 1,700 variables. These were contributed by 41 coders (see https://chield.excd.org/about.html).

The ratio of unique variables to links is high, suggesting that many documents introduce new variables. Ideally, if the database was approaching a "full picture" of the field, the number of new variables being added would decrease over time. Figure 2 shows the relationship between the number of documents and the number of unique variables (points are average number of unique labels from 1000 random orderings of documents). The curve is growing slower than strictly linear, and if we assume a quadratic function, then we estimate that a plateau will be reached when around 650 documents have been coded. Of course, at the moment, the database reflects the research interests of its contributors and might under-represent many sub-fields, so coverage of the whole field might require many more documents. However, another sign of convergence is that documents are highly connected. The largest component of the network includes around 75% of all documents (Figure 3).

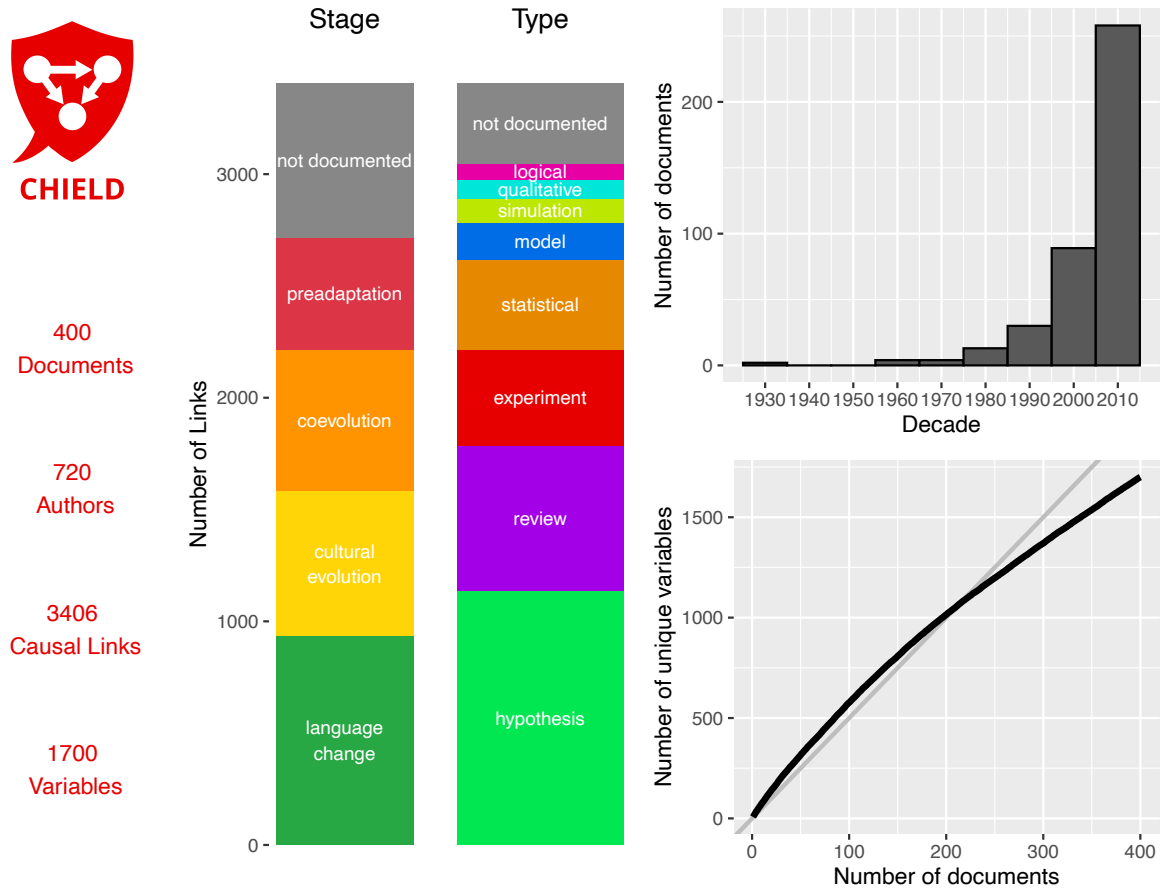

Figure 2: Basic statistics of CHIELD. The first stacked bar shows the number of links for each stage of language evolution. The second stacked bar shows the number of links for each type of evidence. The upper-right panel shows the number of documents by decade of publication. The lower-right panel shows how the number of unique variables grows as documents are added to CHIELD. For context, a linear grey line is shown beneath (intercept = 0, slope = 5).

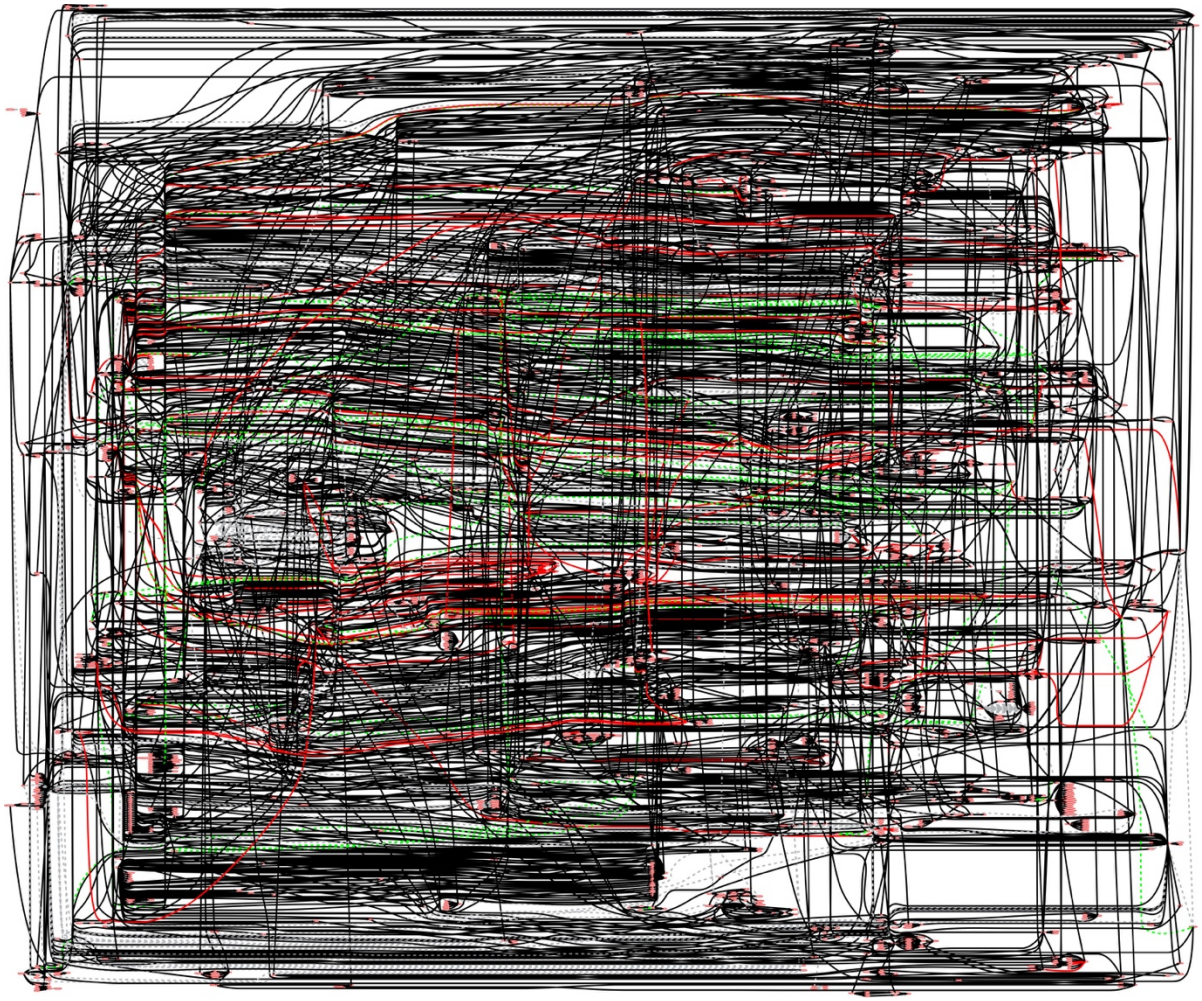

Figure 3: The largest connected component in CHIELD (1,542 variables, 3,222 links).

4.1 Case study 1: gossip, ritual and language

The first example of CHIELD's functionality demonstrates how theories can be compared against each other, and how CHIELD can be used to explore empirical evidence that might help resolve debates. Figure 4 shows two theories about the coevolution of group organisation and communication in human language emergence. The first theory, Dunbar (2004) relates population size, brain size and gossip: risk of predation drives individuals into larger groups for safety, but these groups require more time dedicated to social bonding in order to maintain alliances. Since gossip is more efficient at maintaining a larger number of alliances than one-on-one physical grooming, it was selected for in humans, leading to a coevolution between

population size and brain size (required for gossiping). However, the second theory, Knight, Power & Watts (1995) argues that for gossip to function as a form of social bonding, it needs to be underwritten by another mechanism that guarantees honesty. For them, the value of gossip is socially determined through ritual. Therefore, the first symbolic communications would have been about collaborating in the maintenance of fictions and ritualistic acts which enforce in-group solidarity (e.g., females banding together to conceal signals of menstruation in order to control access to sex, in return for parental investment from males).

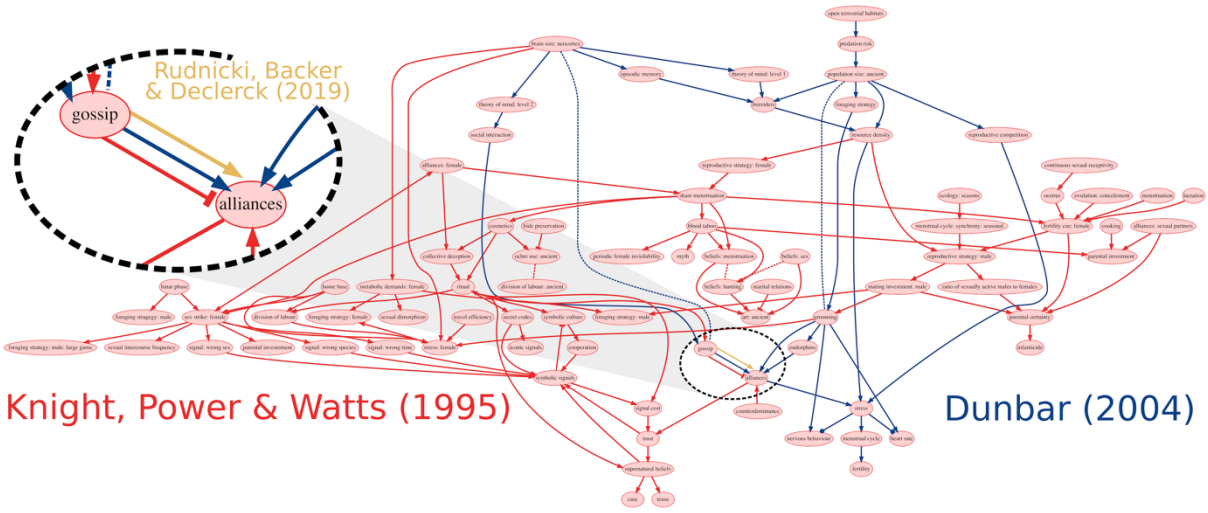

Figure 4: Comparison of Dunbar (2004) and Knight, Power & Watts (1995), with an insert showing the conflict between them and an additional study by Rudnicki, Backer & Declerck (2019) that was automatically discovered.

Both Dunbar's and Knight et al's theories are much more complicated than these simplistic explanations. Nonetheless, they can be suitably represented as causal graphs. Doing so leads to additional evaluative insight: Even though the theories are seen as being totally opposed to each other, when their causal graphs are overlain, there is only one place where they conflict. CHIELD automatically identifies the critical causal link at the heart of the disagreement: whether gossip can maintain alliances (Dunbar argues that it does; Knight, Power and Watts argue that it does not in the absence of ritual bonds). Other parts of the theories are actually mutually compatible.

One of the core strengths of CHIELD is that it can identify studies that test critical causal links. For example, the explore tool automatically discovered an experiment by Rudnicki, Backer & Declerck (2019) where pairs of participants played a trust game after either gossiping for 20 minutes or interacting without gossiping. They found that (for prosocial people) gossiping increased trust, in line with Dunbar's hypothesis. Rudnicki, Backer & Declerck do not discuss Knight, Power & Watts' hypothesis, but the link discovered through CHIELD suggests that their paradigm could be extended to compare the two theories: participants could perform a bonding ritual together rather than gossip. In this way, CHIELD can be an effective tool for identifying critical differences between hypotheses, and also for discovering work that might help resolve the disagreement between them.

4.2 Case study 2: population size and morphological complexity

The next example demonstrates the ability to evaluate multiple different theories. Lupyan & Dale (2010), following theories from studies of language contact, showed that a language's morphological complexity can be predicted by the number of speakers who speak it. They hypothesised that larger populations have more adult learners and more contact with other languages. These factors might cause a pressure for the morphological system of the language to become simpler. For example, adults are worse at learning morphological rules than lexical strategies. That is, languages with large number of speakers might adapt to the adult 'cognitive niche'.

Figure 5 shows the causal graph for this hypothesis, which highlights some key points. First, while the hypothesised mechanism has several steps, the main quantitative result is a correlation between population size and morphological complexity (due to the intervening variables having limited data available). The correlation is consistent with the hypothesis, but alternative data or methods could be applied to try and support each causal link. Secondly, while most links are supported either by reviews from the literature or statistical analyses, there is a "weak link": there was no supporting evidence for a causal effect of population size and the proportion of adult learners. Although it makes logical sense, ideally it should be confirmed empirically (as was recently done in Koplenig, 2019).

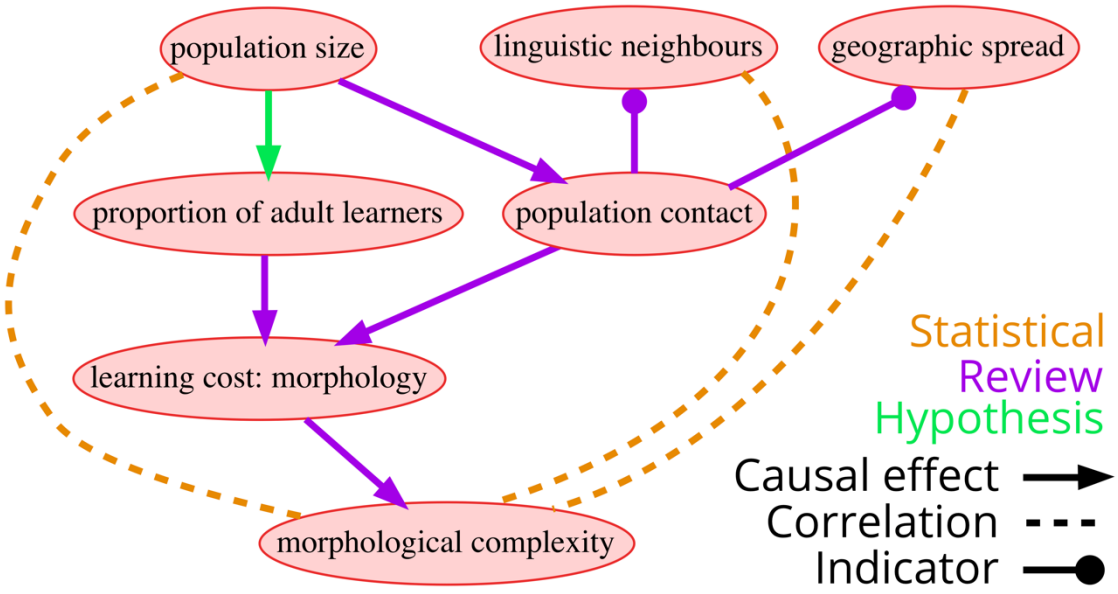

Figure 5: Lupyan & Dale (2010)'s hypothesis, expressed as a causal graph. Links are coloured according to the type of evidence provided.

Lupyan & Dale's study led to several other empirical studies looking at the relationship between morphological complexity and population size, as well as the invocation of previous

studies to explain the patterns. In Figure 6, we show 21 of these studies represented as causal graphs (supporting materials S3 include the R script for automatically generating this figure from CHIELD). This graph includes evidence from fieldwork, cross-cultural statistics, lab experiments and simulations. While this visualisation may look complicated, in tandem with the interactive features of the website it provides a way of systematically thinking about different explanations. For example, Nettle (2012) and Cuskley & Loreto (2016)'s explanation involves general processes, whereby larger populations change frequency distributions in ways that lead to simplification. In contrast, Wray & Grace (2007) and Little (2011) suggest that there are specific effects of the way adults simplify their speech when talking to strangers. Ardell, Anderson & Winter (2016) suggest that the mechanism involves phonetic variation rather than morphology directly, while Atkinson, Kirby & Smith (2015) suggest that phonology is the key.

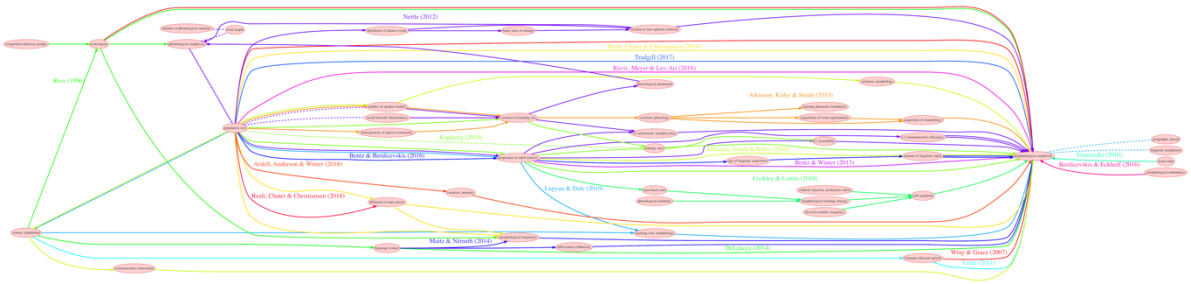

Figure 6: 19 hypotheses for the mechanism that connects population size and morphological complexity.

Interestingly, as the causal graph shows, many of these theories do not necessarily conflict with each other. In fact, there are several nodes in this network that could be explicitly measured in a way that allowed the explanatory power of different causal paths to be tested against each other (e.g., in a causal path regression). Examples of conflicts include the presence or absence of a robust correlation between the proportion of adult learners and morphological complexity (Bentz & Winter, 2013; Koplenig, 2019). Koplenig (2019) finds a robust correlation between population size and complexity in terms of entropy, but not a link to morphological complexity.

The graph also suggests areas where theory might be extended. For example, other factors that influence morphological complexity include word order (Sinnemäki, 2010, see also Koplenig, 2019) and morphological redundancy (Berdicevskis & Eckhoff, 2016). Perhaps these interact with the other variables in a way that might introduce confounds. Following Thurson (1989) and Hymes (1971), Ross (1996) discusses an alternative mechanism - esoterogeny - which predicts that more contact will lead to *greater* morphological complexity, due to competing groups trying to distinguish themselves. It is currently unclear what would be required to test this prediction on a large scale, but some kind of measure of between-group competition would be needed (see Roberts, 2010). These are just some of the ways that causal graph visualisations might inspire future work.

Another contribution to theory that can arise when constructing these graphs, is an understanding of differences in the way terminology is used. For example, while coding some of these studies, it became clear that "population size" was not the only way to refer to how

many speakers a language has. Table 4 shows 10 different terms alongside their definitions. In anthropology, terms are usually more specific in order to capture distinctions, such as the total number of speakers of a language and the number of people in a community (the first may be very high while the second could be fairly low). In contrast, studies in modelling or demography tend to use terms derived from biology that refer to more abstract properties. Dissonance amongst terms is a known issue (at least by experts), but CHIELD provides motivation and a formal system for trying to achieve unification in terminology.

4.3 Case study 3: networks of authors

Figure 7 shows a network of connections between authors, where each connection indicates that causal graphs from hypotheses by the two authors have at least one node in common. This network includes 500 of the 720 authors in the database and 6065 connections. This kind of network is unique, since it is built from hand-coded data about central components of hypotheses rather than publication of co-authorship statistics or textual analysis. Of course, these are based on just a small sample of papers in the field, and heavily biased by the research interests of the coders. Nevertheless, we can use standard network analysis tools like modularity to find clusters of authors. The network splits into three main clusters which might be characterised based on their methods: experimental (experimental semiotics, iterated learning, computational modelling), statistical (cross-cultural statistics, phylogenetics) and comparative (animal communication, genetics). This suggests that researchers mainly cluster on their approaches rather than their topics, meaning that there may be scope for more collaborative work.

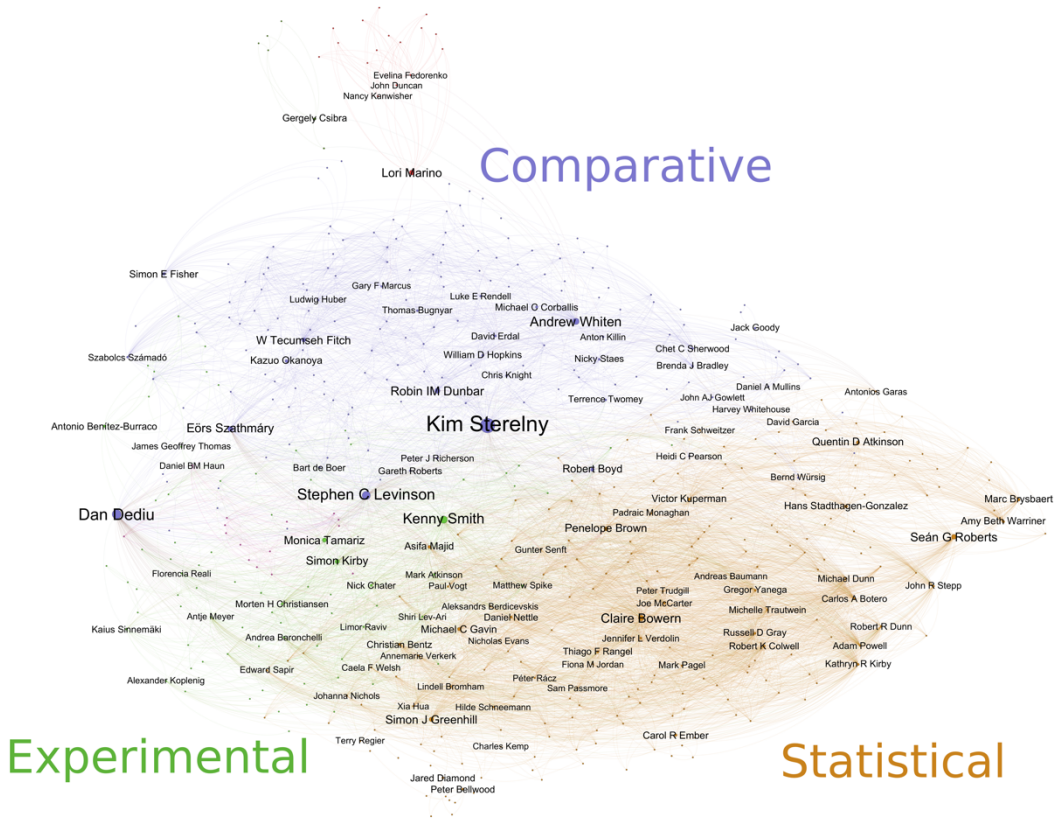

Figure 7: Author network. Connections between authors indicate that causal graphs from hypotheses by the two authors have at least one node in common. The links are coloured by clusters discovered according to network modularity.

Network measures can be used to find ‘brokers’: individuals who provide critical bridges between clusters. For example, betweenness-centrality calculates the shortest path between each pair of nodes, then for each node counts the number of shortest paths that flow through it. This estimates the number of connections that would become longer or potentially break if the given node was not there. Brokers include Kim Sterelny, Stephen Levinson, Kenny Smith, Monica Tamariz, Simon Kirby, Claire Bowern, Simon Greenhill, Dan Dediu and Andrew Whiten. These are generally researchers with interests that span the three main clusters.

This information is used in the online interface for CHIELD to suggest authors who might have interests in common and might make good collaboration partners. Authors are connected in the network if they share a node but have not co-authored a publication (listed in CHIELD). These suggestions are biased by the selection of documents, but generally make sensible suggestions. For example, the top suggested collaborator for Dan Dediu, an executive editor of the Journal of Language Evolution, is Bart de Boer, another executive editor of the same journal.

4.4 Challenges

There were several challenges while building CHIELD, many of which may be applicable to any attempt that catalogues causal theories. Firstly, we found that coding causal graphs from publications is hard. In a few cases, two coders coded the same paper and produced graphs with no overlap. There was even a case where a coder coded a paper twice, several months apart, and the agreement was low. Such problems in coding come from two sources: Firstly, there is often a lack of clarity on the author’s part (which causal graph methods might address). Experiments have shown that researchers can correctly identify causal structures when appropriate information is available (Wiley & Myers, 2003). Secondly, the research priorities of the coder will influence how the graph is coded. For example, biasing the details that they code, or how they choose to divide the elements of the hypothesis into nodes. More specific problems arose regarding the resolution at which to code the paper (general processes versus specific mechanisms), or how to represent structures like trade-offs or interaction effects. While causal links into a node represent a joint conditional probability distribution that does capture the information in an interaction effect, there is no standard way of graphically representing interactions in causal graphs as distinct from simple direct effects (see VanderWeele & Robins, 2007, pp. 1098-1100). However, in many cases the issue can be resolved by thinking about the extra steps in between (see supporting materials S4). We noticed that some coders appeared to have particular ‘styles’, such as aiming to code the hypothesis as a set of discrete steps (where different species have progressed to different extents), or as a dynamic system (where different species have different values at each node). These disagreements lead to practical problems. For example, CHIELD depends on agreement in variable names. Small differences in labelling lead to major changes in the network. Very general node labels such as “language” are also unhelpful when trying to identify causal paths between theories. These issues are mitigated to some extent by the provision of term recommendations on data entry and the ability for anyone to edit causal graphs after initial submission. Fully unifying the vocabulary would take a lot of editing work, but this might be worth the effort in order to improve research communication in the field.

The final challenge was that a theory cannot be fully understood from its graph alone. For this reason, it is strongly encouraged to add quotes from the paper in the "notes" field that include important context. However, combined with some general background knowledge of the field, causal graphs can provide a very helpful overview of ongoing and existing research. Indeed, we found that the process of coding causal graphs increased the coder's own clarity of the paper beyond simply reading the text. We maintain that CHIELD is a useful tool for researchers to organise their research, but we are aware that it would take a huge amount of work to produce a definitive overview of the whole field, if this were at all possible.

The challenges above affect the scalability of the database as the number of contributors and topics expand. We are currently optimistic. Thanks to the use of git and Github, the current system seems adequate for handling the amount of data and edits (around 1000 edits to causal graphs). CHIELD serves a relatively large and diverse set of researchers (41 contributors from multiple disciplines e.g. anthropology, cognitive science, computational modelling, genetics, morphology, philosophy, primatology, psychology, sociolinguistics, syntax, and typology), who have been able to contribute despite many having little technical or programming backgrounds. Although anyone can suggest edits, there is currently only one moderator, and the project may need to expand to one or more committees for dealing with sub-topics or variable names. We suggest that expansion to other fields may be best done by starting a seperate database using the tools and structures of CHIELD as a template (e.g. the Hypothesis Database for Research into the Evolution of Culture, HyDREC https://github.com/j-winters/HyDREC), rather than trying to fit too many fields in a single source. Refining the database may be facilitated by dedicated sessions at workshops. One major bottleneck is training in causal inference methods for the coders. We hope that this can become a more central part of research training in general in the future.

More generally, the data in CHIELD is designed to be used with the Directed Acyclic Graph (DAG) approach to causality and the tools for dealing with them. However, there are limitations on the kinds of causal relations that can be represented in this way, and other approaches are available (see e.g. Mahoney, 2008; Granger, 2016, Blasi & Roberts, 2017). In particular, causal loops can be graphically represented using DAGs, but they violate some of the assumptions that allow various parts of the causal inference machinery to function. Similarly, inference is complicated by non-monotonic effects (VanderWeele & Robins, 2010) or effects happening on different timescales (Aarlen et al. 2016). Many of the theories coded in CHIELD include these features, which limits quantitative treatment and complicates searching for ancestors and links between theories. However, we believe that the current approach still has qualitative value and as methods and theory develop, CHIELD and the data therein can grow to alleviate these shortcomings.

We note that there are existing attempts for automatic extraction of causal relations directly from publication texts (see Alshuwaier, Areshey, & Poon, 2017; Asghar, 2016; Mueller & Abdullaev, 2019; Mueller & Huettemann, 2018; Tshitoyan et al., 2019). However, given the problems above, we are sceptical of the effectiveness of this approach for evolutionary linguistics. In any case, progress in automated coding of causal structure would require human-coded, "gold standard" data, which could be provided by CHIELD.

5 Conclusions

We presented the design and implementation of CHIELD, a database of causal hypotheses in evolutionary linguistics. We demonstrated CHIELD's uses, including identifying critical differences between theories, discovering critical evidence, synthesising theories and finding collaborators. The main challenge is in the coding of hypotheses. However, the challenge derives mainly from the difficulty of accurately conveying causal ideas in prose, rather than any limitation of causal inference tools. This issue would be assisted by the use of causal graphs to express hypotheses in the first place. One possible solution would be for journals to encourage that authors submit a causal graph with each publication (as a figure or as metadata). This could then be more easily fed back into CHIELD order to strengthen the representation of the database.

There are many ways that CHIELD could be extended in the future. For the database itself, we plan to add permanent links to the variables and documents, add statistical results to edges (though there is the question of how to allow multiple types of statistic to one edge), and add more flexibility in searching (e.g., limit the results to a particular stage of language evolution). There is also the possibility of converting to database formats that explicitly support network structures for advanced querying (e.g., GraphML). We would like to add support for large touchscreens and interactive debate so that CHIELD could be used as a tool for live discussion between researchers. As research methods increasingly involve collaboration with multiple researchers, this could help people check how their understanding of terms and concepts match. We would also like to develop a way for CHIELD to automatically suggest studies (e.g., for student projects). For example, randomly choosing a causal link that currently has no empirical evidence, detecting colliders that might require alternative explanations, or showing a link between two theories that might not have been noticed. It would also be possible to link nodes to open source data, so that CHIELD could try to discover links that currently have no empirical support, but for which there was existing data that could be utilised. Of course, automated procedures may be susceptible to bias (e.g., publication biases), and they can be no replacement for careful research practice. But we hope CHIELD can help make research more systematic. In particular, we are keen to explore how the data in CHIELD can be used as prior biases in more automated machine learning processes, such as automatic causal graph inference.

In conclusion, clearly expressing complex scientific hypotheses is hard. There is a need for formal tools, such as causal graphs, to help researchers communicate their ideas. CHIELD provides easier access to these tools and creates a space for clarifying theories. It demonstrates that the field of evolutionary linguistics is more connected than we thought, and there are potentially many links between theories that are waiting to be discovered. We suspect the same is true for many other fields that would also benefit from building their own database of causal graphs.

Acknowledgements

AD, CS, FJ and ETC were funded from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 639291, Starting Grant VARIKIN). JN was supported by a Principal's Career Development Scholarship from the School of Philosophy, Psychology & Language Sciences at Edinburgh. KS was supported by an Academy of Finland grant (296212) and by an ERC starting grant (805371). RS was supported by the Australian Research Council (FL130100111). SGR is supported by a Leverhulme early career fellowship (ECF-2016-435). SFM is supported by an Australian Government Research Training Program (RTP) Scholarship and Australian Research Council Laureate Fellowship Grant FL130100141. CS was supported by the John Templeton Fund (40128). The following researchers contributed data to CHIELD without contributing to this paper: Lachlan Walmsley, Limor Raviv, Andrew Buskell, Cara Evans, Michael Dunn, Robin Dunbar, Isobel Clifford, James Winters and Simon Kirby. Many thanks to Robert Forkel, Tessa Alexander, Jon Hallett and Simon Greenhill for additional code review.

6 References

Aalen, O. O., Røysland, K., Gran, J. M., Kouyos, R., & Lange, T. (2016). Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms. *Statistical methods in medical research*, *25*(5), 2294-2314.

Alshuwaier, F., Areshey, A., & Poon, J. (2017). A comparative study of the current technologies and approaches of relation extraction in biomedical literature using text mining. 2017 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS), 1–13. https://doi.org/10.1109/ICETAS.2017.8277841

Ardell, D., Anderson, N., & Winter, B. (2016). Noise In Phonology Affects Encoding Strategies In Morphology. In S. G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér, & T. Verhoef (Eds.), *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANGX11)*.

Atkinson, M., Kirby, S., & Smith, K. (2015). Speaker input variability does not explain why larger populations have simpler languages. *PloS One*, *10*(6), e0129463.

Asghar, N. (2016). Automatic Extraction of Causal Relations from Natural Language Texts: A Comprehensive Survey. Retrieved April 15, 2019, from undefined website: /paper/Automatic-Extraction-of-Causal-Relations-from-A-Asghar/e115ec4c138cc5157ab24a6e462f1fc74902d53f

Baayen, R. H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. Aphasiology, 30(11), 1174-1220.

Bareinboim E, Pearl J (2012) Causal inference by surrogate experiments: z-identifiability. In de Freitas N. & Murphy K. (eds.) Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence, (AUAI Press, Corvalis, OR), pp 113–120.

Bareinboim, E., & Pearl, J. (2016). Causal inference and the data-fusion problem. Proceedings of the National Academy of Sciences, 113(27), 7345-7352.

Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

Bentz, C., & Winter, B. (2013). Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change*, *3*(1), 1–27.

Berdicevskis, A., & Eckhoff, H. (2016). Redundant Features Are Less Likely To Survive: Empirical Evidence From The Slavic Languages. In S. G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér, & T. Verhoef (Eds.), *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANGX11)*.

Bergmann T. and Dale R. (2016). A Scientometric Analysis Of Evolang: Intersections And Authorships. In S.G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér & T. Verhoef (eds.) The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11). Available online: http://evolang.org/neworleans/papers/182.html

Berwick, R. C., & Chomsky, N. (2013). Birdsong, speech, and language: exploring the evolution of mind and brain. MIT press.

Berwick, R. C., & Chomsky, N. (2016). Why only us: Language and evolution. MIT press.

Blasi, D. E., & Roberts, S. G. (2017). Beyond binary dependencies in language structure. Dependencies in Language, 117-128.

Botha, R. P. (2003). Unravelling the evolution of language. Brill.

Bowern, C. (2015). Linguistics: Evolution and Language Change. Current Biology, 25(1), R41-R43.

Bybee, J. (2015) Language change. Cambridge University Press.

Cangelosi, A., & Parisi, D. (Eds.). (2012). Simulating the evolution of language. Springer Science & Business Media.

Cheney, D. L., & Seyfarth, R. M. (2005). Constraints and preadaptations in the earliest stages of language evolution. The Linguistic Review, 22(2-4), 135-159.

Christiansen, M. H., & Kirby, S. (Eds.). (2003). Language evolution. OUP Oxford.

Coelho, M. T. P., Pereira, E. B., Haynie, H. J., Rangel, T. F., Kavanagh, P., Kirby, K. R., … Gavin, M. C. (2019). Drivers of geographical patterns of North American language diversity. Proceedings of the Royal Society B: Biological Sciences, 286(1899), 20190242.

Corballis, M. C. (1999). The Gestural Origins of Language: Human language may have evolved from manual gestures, which survive today as a" behavioral fossil" coupled to speech. American Scientist, 87(2), 138-145.

Croft, W. (2008) Evolutionary linguistics.Annu. Rev. Anthropol.37,219–234

Culbertson, J., & Newport, E. L. (2015). Harmonic biases in child learners: In support of language universals. Cognition, 139, 71-82.

Currie, A., & Killin, A. (2019). From things to thinking: Cognitive archaeology. Mind & Language, 34(2), 263-279.

Cuskley, C., & Loreto, V. (2016). The Emergence Of Rules And Exceptions In A Population Of Interacting Agents. In S. G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér, & T. Verhoef (Eds.), *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANGX11)*.

Deacon, T.W. (1997) The Symbolic Species: the Co-Evolution of Language and the Brain, Norton

De Boer, B., & Verhoef, T. (2012). Language dynamics in structured form and meaning spaces. *Advances in Complex Systems*, *15*(03n04), 1150021.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. Numerische mathematik, 1(1), 269-271.

Ding, P., & Miratrix, L. W. (2015). To adjust or not to adjust? Sensitivity analysis of M-bias and butterfly-bias. Journal of Causal Inference, 3(1), 41-57.

Dor, D., Knight, C., & Lewis, J. (2014). The social origins of language (Vol. 19). Oxford University Press.

Dunbar, R., & Dunbar, R. I. M. (1998). *Grooming, gossip, and the evolution of language*. Harvard University Press.

Dunbar, R. I. (2004). Gossip in evolutionary perspective. Review of general psychology, 8(2), 100-110.

Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. Nature, 473(7345), 79.

Easterday, M. W., Aleven, V., & Scheines, R. (2007). Tis better to construct than to receive? The effects of diagram tools on causal reasoning. Frontiers in Artificial Intelligence and Applications, 158, 93.

Easterday, M. W., Aleven, V., Scheines, R. & Carver, S. M. (2009). Constructing causal diagrams to learn deliberation. International Journal of Artificial Intelligence in Education, 19 (4), 425-445.

Ellson, J., Gansner, E. R., Koutsofios, E., North, S. C., & Woodhull, G. (2004). Graphviz and dynagraph—static and dynamic graph drawing tools. In Graph drawing software (pp. 127-148). Springer, Berlin, Heidelberg.

Elwert, F. (2013). Graphical causal models. In: Handbook of Causal Analysis for Social Research. S. L. Morgan (ed). Dordrecht: Springer.

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. Annual review of sociology, 40, 31-53.

Falk, D. (2016). Evolution of Brain and Culture. Journal of Anthropological Sciences, 94, 1.

Fehér, O., Ljubicic, I., Suzuki, K., Okanoya, K. & Tchernichovski, O. (2017). Statistical learning in songbirds: from self-tutoring to song culture. Philosophical Transactions of the Royal Society B. 372, 1711.

Fitch, W. T. (2010). The evolution of language. Cambridge University Press.

Gavin, M.C., Botero, C.A., Bowern, C., Colwell, R.K., Dunn, M., Dunn, R.R., Gray, R.D., Kirby, K.R., McCarter, J., Powell, A. and Rangel, T.F. (2013). Toward a mechanistic understanding of linguistic diversity. BioScience, 63(7), pp.524-535.

Goldin-Meadow, S., & Yang, C. (2016). Statistical evidence that a child can create a combinatorial linguistic system without external linguistic input: Implications for language evolution. Neuroscience & Biobehavioral Reviews.

Granger, C. W. (2016). Causal inference. The New Palgrave Dictionary of Economics, 1-4.

Gumperz, J. (1968). The Speech Community. in Duranti, Alessandro (ed.) Linguistic Anthropology: A reader 1:66-73.

Guns, R., & Rousseau, R. (2014). Recommending research collaborations using link prediction and random forest classifiers. Scientometrics, 101(2), 1461-1473.

Hauser, A., & Bühlmann, P. (2012). Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. Journal of Machine Learning Research, 13(Aug), 2409-2464.

von Hardenberg A, Gonzalez-Voyer A (2013). “Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis.” Evolution, 67 - 2, 378-387. doi: 10.1111/j.1558-5646.2012.01790.x.

Heinze-Deml, C., Maathuis, M. H., & Meinshausen, N. (2018). Causal structure learning. Annual Review of Statistics and Its Application, 5, 371-391.

Höfler, M., Venz, J., Trautmann, S., & Miller, R. (2018). Writing a discussion section: how to integrate substantive and statistical expertise. BMC medical research methodology, 18(1), 34.

Hsu, P. S., Van Dyke, M., Chen, Y., & Smith, T. J. (2015). The effect of a graph-oriented computer-assisted project-based learning environment on argumentation skills. Journal of Computer Assisted Learning, 31(1), 32-58.

Hurford (2007) The Origins of Meaning: Language in the Light of Evolution. Oxford University Press.

Hurford, J. (2003). Evolution of language: cognitive preadaptations. In P. Strazny (ed.) Fitzroy Dearborn Encyclopedia of Linguistics, Fitzroy Dearborn Publishers, Chicago.

Hymes, D. H. (Ed.). (1971). *Pidginization and creolization of languages*. CUP Archive.

Irvine, L., Roberts, S., & Kirby, S. (2013). A robustness approach to theory building: A case study of language evolution. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 35, No. 35).

Jon-And, A., & Aguilar, E. (2019). A model of contact-induced language change: Testing the role of second language speakers in the evolution of Mozambican Portuguese. PloS one, 14(4), e0212303.

Jovani, R., & Mavor, R. (2011). Group size versus individual group size frequency distributions: a nontrivial distinction. Animal behaviour, 82(5), 1027-1036.
Kalisch, M., Mächler, M., Colombo, D., Maathuis, M. H., & Bühlmann, P. (2012). Causal inference using graphical models with the R package pcalg. Journal of Statistical Software, 47(11), 1-26.
Kempe, V., Gauvrit, N., & Forsyth, D. (2015). Structure emerges faster during cultural transmission in children than in adults. Cognition, 136, 247-254.
Killworth, P. D., Bernard, H.R., & McCarty, C. (1984). Measuring patterns of acquaintanceship. *Current Anthropology*, *25*(4), 381-397.
Kinsella, A. R. (2009). Language evolution and syntactic theory. Cambridge University Press.
Kirby, S., & Christiansen, M. H. (2003). Language evolution: The hardest problem in science. Language Evolution, 1-15.
Knight, C., Studdert-Kennedy, M., & Hurford, J. (Eds.). (2000). The evolutionary emergence of language: social function and the origins of linguistic form. Cambridge University Press.
Kong, X., Jiang, H., Wang, W., Bekele, T. M., Xu, Z., & Wang, M. (2017). Exploring dynamic research interest and academic influence for scientific collaborator recommendation. Scientometrics, 113(1), 369-385.
Koplenig, A. (2019). Language structure is influenced by the number of speakers but seemingly not by the proportion of non-native speakers. *Royal Society Open Science*, *6*(2), 181274.
Labov, W. (1972). *Sociolinguistic Patterns* . Philadelphia: University of Pennsylvania Press .
Lieberman, 1984; Corballis, 1999; Hurford, 2003; Slocombe & Zuberhühler (2005); Cheney & Seyfarth, 2005; Fehér, 2017; Vernes, 2017;
Lieberman, P. (1984). The biology and evolution of language. Harvard University Press.
Little, H. (2011). *The Role of Foreigner Directed Speech in the Cultural Transmission of Language and the Resulting Effects on Language Typology*.
Lopes, G. R., Moro, M. M., Wives, L. K., & De Oliveira, J. P. M. (2010). Collaboration recommendation on academic social networks. In International conference on conceptual modeling (pp. 190-199). Springer, Berlin, Heidelberg.
Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS one*, *5*(1), e8559.
Mahoney, J. (2008). Toward a unified theory of causality. Comparative Political Studies, 41(4-5), 412-436.
Majid, A., Jordan, F., & Dunn, M. (2015). Semantic systems in closely related languages.
McGrath, Joseph, E. (1984). Groups: Interaction and Performance. Englewood Cliffs, NJ: Prentice-Hall. pp. 61–62.
Middleton, J. A., Scott, M. A., Diakow, R., & Hill, J. L. (2016). Bias amplification and bias unmasking. Political Analysis, 24(3), 307-323.
Morgan, M. S. (2013). Nature's experiments and natural experiments in the social sciences. *Philosophy of the Social Sciences*, *43*(3), 341-357.
Morgan, M. H. (2014). *Speech communities*. Cambridge University Press.
Mueller, R. M., & Huettemann, S. (2018). Extracting Causal Claims from Information Systems Papers with Natural Language Processing for Theory Ontology Learning. https://doi.org/10.24251/HICSS.2018.660

Mueller, R., & Abdullaev, S. (2019). DeepCause: Hypothesis Extraction from Information Systems Papers with Deep Learning for Theory Ontology Learning. Retrieved from http://scholarspace.manoa.hawaii.edu/handle/10125/60059

Mufwene, S. S. (2001). The ecology of language evolution. Cambridge University Press.

Murdock, G. P., & Wilson, S. F. (1972). Settlement patterns and community organization: Cross-cultural codes 3. Ethnology, 11(3), 254-295.

Nettle, D. (2012). Social scale and structural complexity in human languages. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1597), 1829-1836.

Noble, W., & Davidson, I. (1996). Human evolution, language and mind: A psychological and archaeological inquiry. CUP Archive.

Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. Proceedings of the National Academy of Sciences, 96(14), 8028-8033.

Pakendorf, B. (2014). Coevolution of languages and genes. Current opinion in genetics & development, 29, 39-44.

Pearl, J. (2000). Causality: models, reasoning, and inference (Vol. 47). Cambridge Univ Press.

Pearl, Judea, & Mackenzie, D. (2018). The Book of Why: The New Science of Cause and Effect. Allen Lane.

Piantadosi, S. T., & Fedorenko, E. (2017). Infinitely productive language can arise from chance under communicative pressure. Journal of Language Evolution, 2(2), 141-147.

Power, C., Finnegan, M., & Callan, H. (2016). Human Origins: Contributions from Social Anthropology.

Progovac, L. (2015). Evolutionary syntax. Oxford University Press.

Progovac, L. (2019). A Critical Introduction to Language Evolution: Current Controversies and Future Prospects. Springer International Publishing.

Pyers, J. E., Shusterman, A., Senghas, A., Spelke, E. S., & Emmorey, K. (2010). Evidence from an emerging sign language reveals that language supports spatial cognition. *Proceedings of the National Academy of Sciences*, *107*(27), 12116-12120.

Ritt, N. (2004). Selfish sounds and linguistic evolution: A Darwinian approach to language change. Cambridge University Press.

Roberts, G. (2010). An experimental study of social selection and frequency of interaction in linguistic diversity. *Interaction Studies*, *11*(1), 138-159.

Roberts, S. G. (2018). Robust, causal, and incremental approaches to investigating linguistic adaptation. Frontiers in psychology, 9, 166.

Roberts, S. G., & Winters, J. (2013). Linguistic diversity and traffic accidents: Lessons from statistical studies of cultural traits. PloS one, 8(8), e70902.

Rogerson, P. A. (1997). Estimating the size of social networks. *Geographical Analysis*, *29*(1), 50-63.

Rohrer, J. M. (2018). Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. Advances in Methods and Practices in Psychological Science, 1(1), 27–42.

Ross, M. D. (1996). Contact-induced Change and the Comparative. In M. Durie & M. D. Ross (Eds.) The comparative method reviewed: Regularity and irregularity in language change, p. 180. Oxford University Press on Demand.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36.

Sampson, G., Gil, D., & Trudgill, P. (Eds.). (2009). Language complexity as an evolving variable (Vol. 13). Oxford University Press.
Sampson, G., Gil, D., & Trudgill, P. (Eds.). (2009). Language complexity as an evolving variable (Vol. 13). Oxford University Press.
Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. Trends in Cognitive Sciences, 14(9), 411–417.
Sinnemäki, Kaius (2010). Word order in zero-marking languages. Studies in Language 34(4), 869–912.
Slocombe, K. E., & Zuberbühler, K. (2005). Functionally referential communication in a chimpanzee. Current Biology, 15(19), 1779-1784.
Smith, K. and Kirby, S. (2008) Cultural evolution: implications for understanding the human language faculty and its evolution. Philosophical Transactions of the Royal Society B: Biological Sciences. 363, 3597–3603.
Steels, L. (1997). The synthetic modeling of language origins. Evolution of communication, 1(1), 1-34.
Tallerman, M. (2007). Did our ancestors speak a holistic protolanguage?. Lingua, 117(3), 579-604.
Tamariz, M., & Kirby, S. (2016). The cultural evolution of language. Current Opinion in Psychology, 8, 37-43.
Textor, J., Hardt, J., & Knüppel, S. (2011). DAGitty: a graphical tool for analyzing causal diagrams. Epidemiology, 22(5), 745.
Thurston, W. (1989). How exoteric languages build a lexicon: Esoterogeny in Western New Britain. In: Ray Harlow, Robin Hooper (eds.) 1989: VICAL 1: Oceanic Languages, Papers from the Fifth International Conference on Austronesian Linguistics. Auckland: Linguistic Society of New Zealand, 555–579.
Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., Persson, K.A., Ceder, G. and Jain, A., 2019. Unsupervised word embeddings capture latent knowledge from materials science literature. Nature, 571(7763), p.95.
Tubau, E. (2008). Enhancing probabilistic reasoning: The role of causal graphs, statistical format and numerical skills. Learning and Individual Differences, 18(2), 187-196.
VanderWeele, T. J., & Robins, J. M. (2007). Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect. American journal of epidemiology, 166(9), 1096-1104.
VanderWeele, T. J., & Robins, J. M. (2010). Signed directed acyclic graphs for causal inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(1), 111-127.
van der Bijl, W. (2018). “phylopath: Easy phylogenetic path analysis in R.” PeerJ, 6, e4718. doi: 10.7717/peerj.4718, R package version 1.0.2.
Vernes, S. C. (2017). What bats have to say about speech and language. Psychonomic bulletin & review, 24(1), 111-117.
Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution.
Vogt, P., & De Boer, B. (2010). Editorial: Language evolution: computer models for empirical data. Adaptive Behavior-Animals, Animats, Software Agents, Robots, Adaptive Systems, 18(1), 5–11.
Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. PloS one, 11(3), e0152719.

Wiley, J., & Myers, J. L. (2003). Availability and Accessibility of Information and Causal Inferences From Scientific Text. Discourse Processes, 36(2), 109–129. https://doi.org/10.1207/S15326950DP3602_2

Woodward, J. (2003). Making things happen: A theory of causal explanation. Oxford: Oxford University Press.

Woodward, J. (2016). Causation and manipulability. In: Stanford encyclopedia of philosophy (winter 2016 edition). E. N. Zalta (ed). https://plato.stanford.edu/archives/win2016/entries/causation-mani/

Wray, A. (2002). Transition to Language. Oxford University Press.

Wray, A., & Grace, G. W. (2007). The consequences of talking to strangers: evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, (117).

Xu, Y., Chang, Z., Lin, J., Ma, J., Hao, J., & Zhao, D. (2010, May). Network based approach for discovering academic researchers with shared interests. In 2010 International Conference on E-Business and E-Government (pp. 1864-1867). IEEE.

Yan, E., & Guns, R. (2014). Predicting and recommending collaborations: An author-, institution-, and country-level analysis. Journal of Informetrics, 8(2), 295-309.

York, R. (2018). Control variables and causal inference: a question of balance. International Journal of Social Research Methodology, 21(6), 675-684.

Youngblood, M., & Lahti, D. (2018). A bibliometric analysis of the interdisciplinary field of cultural evolution. Palgrave Communications, 4(1), 120.

Zuidema, W. & de Boer, B. (2010) Models of Language Evolution: Does the Math Add Up? Models of Language Evolution Workshop at EVOLANG 2010, April 14, 2010, Utrecht, the Netherlands

Zuidema, W. & de Boer, B. (2013) Modeling in the language sciences. In: Podesva, R. J. & Sharma, D. (Eds.) *Research Methods in Linguistics.* Cambridge: Cambridge University Press, pp. 428–445.